



Facultad de Ciencias

**Segmentación de gliomas en Imagen de
Resonancia Magnética Multimodal
(Glioma segmentation in Multimodal Magnetic
Resonance Imaging)**

Trabajo de Fin de Máster
para acceder al

MÁSTER EN DATA SCIENCE

Autor: Andrea Noriega Puente

**Director: David Rodríguez González
Codirectora: Lara Lloret Iglesias**

Octubre - 2019

Agradecimientos

En primer lugar quisiera mostrar mi más sincero agradecimiento a David Rodríguez, mi director de TFM, y a Lara Lloret, mi codirectora, ambos además profesores durante el Máster. En primer lugar, por estar siempre disponibles para solventar cualquier duda que me surgiera, ya sea en el despacho o por Skype, y por responder a mis correos casi antes de haberlos enviado. Gracias por calmar mis agobios cuando no obtenía resultados y por guiarme por el buen camino, han sido un pilar fundamental en su desarrollo, y siempre han estado ahí cuando he necesitado cualquier empujón.

En segundo lugar, a mis padres, Conchi y Ricardo, y a Hugo, que aunque no tuvieran mucha idea de lo que les contaba sobre el Trabajo, siempre han tenido un momento para escuchar mis avances o explicaciones imposibles. Gracias también por aguantar mis quejas y agobios, y por animarme a seguir trabajando duro cuando el código no hacía otra cosa que dar errores.

Y por último quiero agradecer a todas aquellas personas que de una forma u otra se han ido interesando por el avance de mi proyecto, y que siempre se han alegrado de los progresos que he ido haciendo.

Resumen

El glioma es el tipo de tumor cerebral más común, presentando distintos grados de malignidad y agresividad, así como un pronóstico variable. La gran variabilidad que caracteriza a estas lesiones implica un estudio completamente único, así como un tratamiento personalizado. Es por esto que la segmentación manual de este tipo de lesiones supone tanto un gran consumo de tiempo como un trabajo duro para el experto, el cual implica el análisis multimodal de varias secuencias diferentes, normalmente, de Resonancia Magnética.

En este trabajo se presenta un modelo de segmentación automática de gliomas de alto grado, HGG, cuya arquitectura se basa en el empleo de redes neuronales convolucionales simétricas, denominada U-Net. Para el entrenamiento del modelo se ha hecho uso del conjunto de datos del Data Challenge BraTS 2018 (Brain Tumor Segmentation), compuesto por un total de 237 estudios de IRM, 162 HGG y 75 LGG, de los que solo se han empleado los primeros. Los resultados en cuanto a la obtención de mapas binarios de segmentación de los tumores han sido positivos. Se ha calculado el coeficiente DICE medio, relacionado con la calidad de la determinación de la forma del tumor, logrando 0.64 ± 0.39 para el modelo entrenado con imágenes T1, y 0.57 ± 0.41 para el modelo entrenado con secuencias T1 con contraste. Cualitativamente se ha observado además un buen desempeño a la hora de localizar espacialmente los tumores, incluso si la forma no es determinada tan precisamente.

Palabras clave: Deep Learning, glioma, segmentación, neuroimagen, imagen médica.

Abstract

Glioma is the most common type of brain tumor, having different degrees of malignancy and aggressiveness, as well as a variable prognosis. The great variability of these pathologies implies a completely unique study, as well as a personalized treatment. This is why the manual segmentation of brain tumors by an expert entails both a big time consumption and a really hard work, involving the multimodal analysis of several different sequences of Magnetic Resonance.

In this project, an automatic segmentation model of high-grade glioma, HGG, is introduced, whose architecture is based in symmetric convolutional neural networks, called U-Net. For training the model, data from the Data Challenge BraTS 2018 (Brain Tumor Segmentation) have been used, consisting of a total of 237 MRI studies, 162 HGG cases and 75 LGG cases, from which only the first ones have been used. Resulting binary segmentation map of the tumors have been positive. The mean DICE coefficient, related to the quality of tumor shape determination, has been calculated, achieving 0.64 ± 0.39 for the model trained with T1 sequences, and 0.57 ± 0.41 for the model trained with T1 enhanced sequences. Qualitatively, good performance has also been observed when determining spatial location of the tumors, even if the shape is not determined so precisely.

Key words: Deep Learning, glioma, segmentation, neuroimaging, medical image.

Índice

Agradecimientos	I
Resumen	III
Abstract	V
1. Introducción	1
1.1. El cerebro humano	1
1.2. Los tumores cerebrales	3
1.2.1. Clasificación	3
1.2.2. Los gliomas	4
1.2.3. Diagnóstico	4
1.2.4. Tratamiento	6
1.3. Funcionamiento y conceptos básicos de la IRM	7
1.4. Motivación	9
2. Estudio del estado del arte	11
2.1. Evolución del análisis de imagen médica	11
2.2. Usos del Deep Learning en análisis de imagen médica	12
2.3. Algoritmos de Deep Learning para segmentación de tumores cerebrales	13
2.4. BraTS challenge	14
3. Dataset	15
3.1. Algunos repositorios de imagen de cerebro	15
3.2. Dataset BraTS 2018	16
3.2.1. Descripción del dataset de entrenamiento	16
3.2.2. Descripción del dataset de validación	18
3.2.3. Formato de las imágenes	18

4. Métodos	21
4.1. Librerías de Python	21
4.1.1. NiBabel	21
4.1.2. Keras	21
4.2. Preprocesamiento de las imágenes	22
4.3. Arquitectura de la red	24
4.4. Definición del modelo y entrenamiento de la red	25
4.5. Predicciones	26
5. Resultados y análisis	27
5.1. Introducción	27
5.2. Resultados con datos de Training	28
5.2.1. Primeros resultados	28
5.2.2. Resultados definitivos	28
5.3. Resultados con datos de Validation	29
6. Discusión	31
7. Conclusiones y trabajo futuro	35
7.1. Conclusiones	35
7.2. Trabajo futuro	36
Bibliografía	42
A. Primeros resultados	43
B. Resultados definitivos	45

Índice de figuras

1.1. Anatomía del sistema nervioso central	2
1.2. Neuronas y células gliales	3
1.3. Imagen del cerebro de una mujer de 56 años con oligodendrioma	5
1.4. Principios básicos de la Imagen por Resonancia Magnética	7
1.5. Comparación secuencias T1 y T2	9
3.1. Secuencias de RM del cerebro de un paciente	17
3.2. Segmentación manual de una imagen de RM	17
4.1. Secuencia de tres rebanadas de cerebro casi consecutivas, con sus correspondientes etiquetas	23
4.2. Proceso de binarización de la etiqueta de una imagen de RM.	23
4.3. Arquitectura de U-Net	24
5.1. Predicciones set Validation	29
6.1. Binarización de las predicciones	32
A.1. Primeras predicciones	44
B.1. Primeras predicciones T1 tumor	46
B.2. Primeras predicciones T1 tumor	47

Capítulo 1

Introducción

En el presente Capítulo se introducirán los conceptos relacionados con la anatomía y fisiología del cerebro humano, así como de la fisiopatología de los tumores cerebrales, los métodos de diagnóstico y tratamientos actuales. Por otro lado se expondrá la motivación para realizar el trabajo.

1.1. El cerebro humano

El sistema nervioso está formado por el sistema nervioso central (SNC) y el periférico (SNP). El primero de ellos está constituido por el cerebro y la médula espinal, ambos protegidos por estructuras óseas, que son el cráneo y la columna vertebral, ver Figura 1.1. El cerebro es una masa de tejido blando protegido por los huesos del cráneo y unas membranas llamadas meninges, entre las cuales fluye el líquido cefalorraquídeo a través de los ventrículos [1]. A su vez, el cerebro está compuesto por un trillón de células aproximadamente, unos 100 billones de neuronas y, aunque su peso constituye solo el 2 % del peso total corporal (alrededor de 1.4 kg), recibe el 20 % del riego sanguíneo total [2].

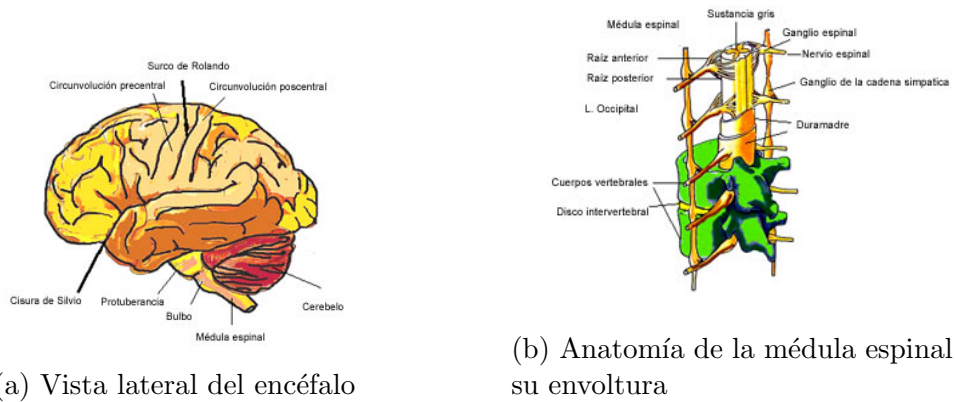


Figura 1.1: Anatomía del sistema nervioso central. Fuente: [3]

Como complemento de las neuronas, se encuentran otros componentes del sistema nervioso, las células gliales. Estas desempeñan funciones auxiliares del tejido nervioso, y constituyen una red intraneural en la que hay una gran variedad de células estrelladas y fusiformes (forma de huso, elipsoidales).

Las células gliales, a diferencia de las neuronas, no establecen conexiones sinápticas, y su principal función es la de asegurar el mantenimiento del equilibrio de las neuronas, es decir, hacen de soporte de estas. Además, se encargan de mantener las condiciones homeostáticas (niveles adecuados de oxígeno y nutrientes), de regular las funciones metabólicas y de hacer de aislante en los tejidos nerviosos, dado que conforman las vainas de mielina que protegen y aíslan los axones de las neuronas. Por este último motivo se dice que las células gliales desempeñan un rol fundamental en la comunicación neuronal.

Dentro del grupo de células gliales diferenciamos entre las centrales, constituyentes del SNC, y las periféricas, constituyentes del SNP. Dentro del primer grupo se encuadran, por ejemplo, los astrocitos, oligodendrocitos y la microglía (ver Figura 1.2), mientras que en el segundo grupo se incluyen las células de Schwann y las de Müller [4].

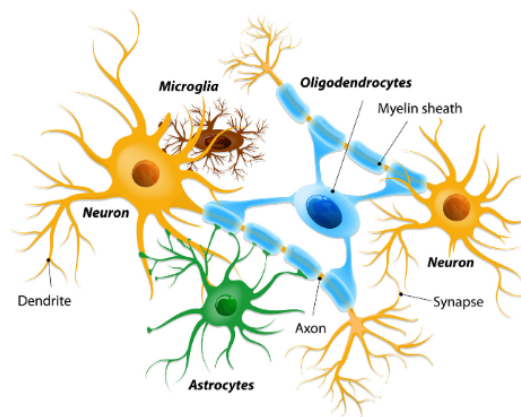


Figura 1.2: Neuronas junto a diferentes tipos de células gliales, las cuales hacen de soporte de las neuronas y conforman las vainas de mielina de los axones.

1.2. Los tumores cerebrales

En general, se denomina tumor cerebral a la presencia de una "masa" que crece dentro del cerebro, siendo primario cuando se origina en el propio cerebro, y secundario o metastásico si se origina en otra parte del organismo y en algún momento de su evolución se extiende al cerebro [1]. Una característica típica de los tumores cerebrales es que, a diferencia de otros tumores malignos de otras localizaciones, rara vez se diseminan fuera del sistema nervioso central [5]. En esta sección se darán detalles básicos en cuanto a la clasificación, diagnóstico y tratamiento de estos tumores.

1.2.1. Clasificación

Para establecer los tipos histológicos, es decir, las variedades de tumores cerebrales, hay que basarse en dos aspectos esenciales. Por un lado, el tipo de célula en la que se origina el tumor, de la cual tomará el nombre. Y por otro lado, se tiene en cuenta el grado histológico o la agresividad del tumor, que se evalúa desde el grado I hasta el IV.

En este segundo tipo de clasificación, el tumor de grado I se considera benigno y de crecimiento lento. El de grado II también es de crecimiento lento, pero podría evolucionar con el tiempo a uno de grado mayor. Y por último, los tumores de grados III y IV son malignos, siendo los de grado IV los más agresivos y con un crecimiento muy rápido [6, 7].

1.2.2. Los gliomas

Los gliomas son un tipo de cáncer poco común, ya que afecta a 6 de cada 100000 personas al año. Sin embargo, representa el 80% de los tumores del Sistema Nervioso Central. Estos tumores pueden afectar a personas de todas las edades, siendo más frecuente en el rango de 50 a 60 años.

Según el tipo de célula en el que se originan, los gliomas pueden ser, por ejemplo, astrocitomas, oligodendrogliomas y oligoastrocitomas, porque tienen su origen en astrocitos, oligodendrocitos o en ambos, respectivamente.

Según el grado de malignidad, los gliomas de grados I y II se denominan de bajo grado (LGG, Low Grade Glioma), mientras que los gliomas de grados III, glioma anaplásico, y IV, glioblastoma, son de alto grado (HGG, High Grade Glioma) [8].

1.2.3. Diagnóstico

A la hora de detectar un tumor cerebral, deben hacerse varias pruebas que, en combinación, permitan determinar las características del mismo (grado de actividad, extensión, etc.). El primer paso es realizar una exploración, tanto física como neurológica, para determinar las siguientes pruebas a realizar. Dentro de estas se encuadran varias técnicas, descritas a continuación [1].

- **Tomografía Axial Computerizada (TAC).** A partir de la emisión de Rayos X se obtienen imágenes de distintas secciones de la cabeza, pudiendo así localizar el tumor y distinguir sus características con gran resolución. Es común la administración de un contraste para mejorar la visualización de ciertas estructuras.
- **Resonancia Magnética (RM).** Es la prueba diagnóstica de primera elección a la hora de detectar tumores cerebrales, ya que permite la obtención de imágenes más precisas del tumor, tanto en cuanto a sus características como a su localización. Suele administrarse también un contraste, pero diferente al del TAC. Su funcionamiento será explicado más detalladamente en la Sección 1.3.
- **Tomografía por Emisión de Positrones (PET).** A diferencia de la RM, no aporta información anatómica detallada, pero puede usarse como complemento de las técnicas mencionadas anteriormente ya que aporta información metabólica. Precisa de la administración al paciente de un radiofármaco, o trazador, que dependiendo de las células que se quieran detectar, se compone de una molécula determinada a la que se

adhiera un isótopo radiactivo. El PET muestra por tanto la distribución regional de la concentración del trazador. Por ejemplo, en oncología se emplea 2-(^{18}F)-fluoro-2-deoxi-Dglucosa (18FDG), compuesto por glucosa y Fluor-18, ya que su concentración en las células tumorales es un reflejo del aumento de su metabolismo glicídico para poder mantener una elevada tasa de crecimiento y proliferación [9].

- **Tomografía por Emisión de Fotón Único (SPECT).** También precisa de la administración de un radiotrazador, y también se emplea como complemento de otras técnicas, ya que permite un análisis más funcional. Es de gran utilidad en el seguimiento de tumores y en la identificación del grado de malignidad.
- **Punción lumbar.** Consiste en la toma de una muestra de líquido cefalorraquídeo para buscar células tumorales o signos de infección.
- **Biopsia.** Es una prueba esencial, básica y necesaria para efectuar el diagnóstico definitivo. Consiste en realizar un examen histopatológico, en el que una muestra extirpada del tumor es analizada al microscopio. De este modo es posible conocer de qué tipo de tumor se trata, qué tipo de células lo componen y cuál es su agresividad.

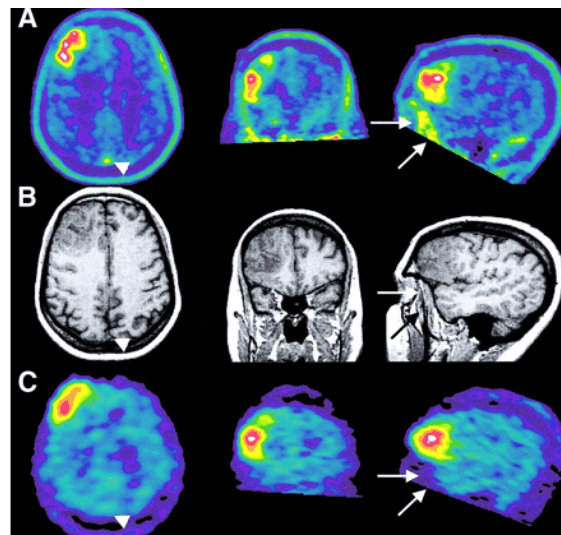


Figura 1.3: Imagen del cerebro de una mujer de 56 años con oligodendrioma. Se muestran secuencias de PET (A), RM (B) y SPECT(C) [10].

1.2.4. Tratamiento

Para combatir el tumor cerebral, el paciente deberá someterse a un tratamiento para aliviar síntomas tales como dolores de cabeza, crisis convulsivas, pérdida de funciones motoras y/o sensitivas, etc. Además, también deberá planificarse una terapia específica, que constará de cirugía con o sin radioterapia, o solo radioterapia en caso de no poderse realizar la cirugía. Normalmente, ambas intervenciones suelen complementarse con la quimioterapia [1].

- **Cirugía.** Es la primera y la principal herramienta para el tratamiento de la mayoría de los tumores. A veces, este paso es suficiente para controlar la enfermedad.
- **Radioterapia.** Es otra herramienta fundamental. Se basa en el envío de partículas ionizadas de gran energía contra las células del tumor, con el objetivo de producir daños en su material genético, facilitando así la muerte celular del tumor. Dado que se trata de radiación ionizante que puede dañar el tejido sano que rodea al tumor, lo habitual es fraccionar la dosis aplicada en varios días o semanas.
- **Quimioterapia.** Históricamente se denomina así al conjunto de medicamentos empleados para el tratamiento específico del cáncer. Esta técnica ha venido presentando dos problemas a la hora de tratar tumores cerebrales: por un lado, la barrera hematoencefálica (BHE), que protege al SNC de la llegada de sustancias tóxicas por la sangre. Y por otro lado, la resistencia histórica de estos tumores al tratamiento antitumoral.

Sin embargo, hoy en día existe otro tipo de medicamentos, llamados biológicos, que no son quimioterapia como tal, ya que han sido diseñados a partir de nuevos conocimientos moleculares y con mecanismos de acción diferentes y más selectivos que la quimioterapia. En general, estos medicamentos biológicos actúan contra una diana previamente determinada, lo que se denomina medicina personalizada. El tratamiento suele combinar quimioterapia con medicamentos biológicos, ya que en muchos casos los resultados obtenidos son mejores que solo con quimioterapia. Un ejemplo de medicamentos biológicos es la inmunoterapia, la cual ha demostrado recientemente obtener resultados satisfactorios contra el glioblastoma [11].

1.3. Funcionamiento y conceptos básicos de la IRM

La Imagen por Resonancia Magnética (IRM) es una técnica no invasiva que permite obtener imágenes anatómicas tridimensionales con un gran nivel de detalle. Se denomina no invasiva puesto que no emplea radiación ionizante ni dañina para el paciente, a diferencia del TAC. En su lugar, esta técnica se basa en el empleo de un potente campo magnético estático, generado por grandes imanes, que hace que los protones del agua que compone los tejidos vivos se alineen. Este proceso, descrito a continuación, se ilustra en la Figura 1.4.

Al emitir un pulso de radiofrecuencia a través de un paciente, los protones son estimulados y giran fuera del equilibrio, luchando contra el campo. Cuando este pulso cesa, los sensores de IRM detectan la energía liberada mientras los protones se realinean con el campo magnético. El tiempo que tardan en volver al equilibrio, así como la cantidad de energía liberada, dependen del entorno y de la naturaleza química de las moléculas [12].

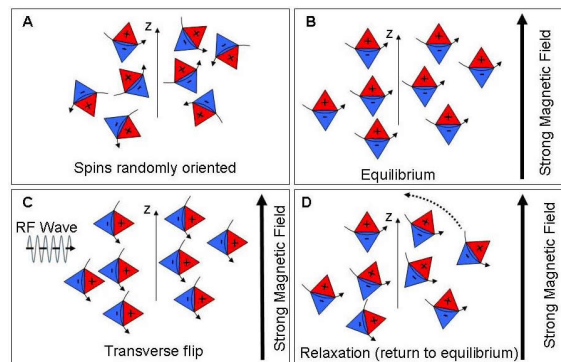


Figura 1.4: Principios básicos de la Imagen por Resonancia Magnética. A: Los espines de los protones están orientados aleatoriamente. B: Los espines se orientan según la dirección del campo magnético aplicado. C: El pulso de radiofrecuencia obliga a los protones a rotar en contra del campo magnético. D: Cuando el pulso de radiofrecuencia cesa, los espines vuelven a la alineación inicial. Fuente: [13]

Cuando dicho pulso se detiene, los protones tienen una cierta magnetización longitudinal, mientras que la componente transversal es nula debido a que pierden la fase. El tiempo que tardan en recuperar el 63% de cada una de las componentes se denominan, respectivamente, T1 y T2, siendo T1 siempre mayor que T2 [14].

En un escáner de Resonancia Magnética, los píxeles de la imagen resultante se pueden ponderar respecto a T1 o a T2. Esto se consigue variando los factores extrínsecos, es decir, los parámetros, del escáner, para que el valor del píxel dependa, sobre todo, de T1 o T2. Dependiendo del tejido que se observe, este puede verse más intenso (más blanco), denominándose hiperintenso, o menos intenso (más negro), denominándose hipointenso [15]. Cuando el tejido tiene tonos similares a tejidos vecinos se denomina isointenso. Dependiendo del tipo de tejido, este será hipo o hiperintenso en función de la ponderación de la imagen:

- **Ponderada en T1.** Si el tejido tiene T1 elevado, será hipointenso, viéndose más oscuro en la imagen.
- **Ponderada en T2.** Si el tejido tiene T2 elevado, será hiperintenso, viéndose más brillante en la imagen.

El retorno de los espines al equilibrio ocurre con diferentes velocidades dependiendo del tejido en que se encuentren los protones. Esta diferencia en la velocidad de relajación se traduce en contraste entre los diferentes tejidos, lo cual permite diferenciarlos en una imagen de Resonancia Magnética. En la Tabla 1.1 se recogen diferentes tejidos observados en IMR, y cómo se ven dependiendo de si la imagen es potenciada en T1 o T2 [15].

Al emplear un agente de contraste en RM, lo que se analiza es el efecto paramagnético que este produce sobre los protones adyacentes, ya que sus tiempos de relajación T1 y T2 se ven modificados.

El gadolinio es un metal del grupo de los lantánidos que tiene electrones impares, por lo que tiene un alto poder paramagnético. Esto implica que la intensidad del campo magnético a su alrededor aumenta, facilitando la relajación longitudinal de los protones vecinos, y su T1 se acortará. Por tanto, se observará hiperintensidad de señal en las secuencias ponderadas en T1, aumentando las diferencias de intensidad de señal entre tejido vascularizados y no vascularizados, y mejorando la capacidad diagnóstica de la Resonancia Magnética [16].

Tejido	T1	T2	FLAIR
Gas (aire)	hipo	hipo	hipo
Hueso	hipo	hipo	hipo
Agua	hipo	hiper	hipo
Agua con proteínas (moco)	depende	depende	depende
Grasa	hiper	hipo	hipo
Sustancia blanca	levemente hiper	hipo	hipo
Sustancia gris	iso	iso	iso
Líquido cefalorraquídeo	hipo	hiper	hipo
Sangre	depende	depende	depende

Tabla 1.1: Cómo se observan diferentes tejidos según la potenciación de la imagen. El agua con proteínas y la sangre son casos particulares. En el primer caso, la concentración de proteína determinará si la sustancia acuosa se ve hiper o hipo . Por otro lado, la sangre, al contener hierro y otras sustancias, se verá diferente dependiendo del estadio de oxidación en el que se encuentre.



(a) Secuencia ponderada en T1 (b) Secuencia ponderada en T2

Figura 1.5: Comparación de dos secuencias de cerebro, una potenciada en T1 (a) y la otra en T2 (b). Las flechas rojas señalan a una zona que contiene agua, viéndose hipointenso en T1 e hiperintenso en T2, como bien se ha descrito en la Tabla 1.1. Fuente: [15].

1.4. Motivación

El glioma es el tipo más común de tumor cerebral, con diferentes grados de malignidad y agresividad, así como pronóstico variable y distintas regiones histológicas. Esta heterogeneidad intrínseca de los gliomas se aplica también

a su apariencia y forma, lo cual se refleja en intensidades variables en los escáneres de resonancia magnética, y que se traduce en propiedades biológicas variables del tumor [17].

Actualmente el diagnóstico de este tipo de patologías se realiza manualmente mediante métodos tanto cualitativos como cuantitativos. Sin embargo, esta tarea tiene una gran dificultad y requiere de mucho tiempo para ser llevada a cabo. Por ejemplo, un estudio multimodal de Imagen de Resonancia Magnética para un glioma, en el cual se analizan varias secuencias diferentes, requiere alrededor de 60 minutos.

Este es el motivo de que la segmentación de tumores cerebrales en imagen de resonancia magnética sea hoy en día uno de los principales desafíos del campo del análisis de imagen médica.

Es por esto que surge una necesidad de implementar algoritmos que puedan llevar a cabo la tarea de segmentación de manera automática, evitando además la variabilidad intrínseca del observador.

De este modo, se conseguiría desarrollar un método de ayuda al radiólogo, llevando a cabo una primera identificación y caracterización biológica del tumor [7].

Dada su elevada relevancia clínica, así como su naturaleza desafiante, parece razonable que el problema de la segmentación de tumores cerebrales haya estado en el punto de mira durante los últimos 20 años. Esto ha dado como resultado un amplio abanico de algoritmos de segmentación automática, semiautomática e interactiva de las diferentes estructuras de los tumores [18], lo que añade un gran punto a favor del interés por realizar este Trabajo de Fin de Máster.

Capítulo 2

Estudio del estado del arte

En este Capítulo se da una visión genérica del estado del arte en cuanto a técnicas de análisis de imagen médica, para profundizar posteriormente en la segmentación de tumores cerebrales.

2.1. Evolución del análisis de imagen médica

Tan pronto como fue posible obtener y guardar imágenes médicas en un ordenador, se han ido desarrollando sistemas que permiten el análisis automático de las mismas.

Desde los años 70 hasta los 90, el análisis de imagen médica era realizado mediante la aplicación secuencial de un modelo de procesamiento de píxeles de bajo nivel, junto con modelos matemáticos, para construir componentes basados en reglas para resolver ciertas tareas en concreto.

A finales de los años 90, los métodos supervisados, entrenados para desarrollar sistemas de aprendizaje automático, fueron ganando cada vez más popularidad entre los métodos de análisis de imagen médica.

A día de hoy las técnicas de Machine Learning para el reconocimiento de patrones constituyen la base de muchos de los sistemas de análisis de imagen médica más exitosos del mercado. De este modo, se ha observado la evolución de sistemas diseñados en su totalidad por humanos hacia sistemas entrenados por ordenadores empleando datos de muestra, de los cuales se obtienen vectores de características. Este último paso de extracción de características es realizado en la mayoría de los casos por los investigadores. Sin embargo, los sistemas avanzan rápidamente hacia el aprendizaje de las características directamente de los datos [19].

2.2. Usos del Deep Learning en análisis de imagen médica

Como bien es sabido, hoy en día la aplicación de técnicas de deep learning en diferentes ámbitos está en continuo crecimiento. Más concretamente, en estudios donde se trata con imágenes, arquitecturas como las redes neuronales convolucionales o recurrentes se han hecho un amplio hueco. A continuación se describen las principales aplicaciones de estas técnicas [19]:

- Clasificación, tanto de imágenes completas, como de partes aisladas dentro de estas. El primer tipo correspondería a la clasificación de una imagen como positivo o negativo en una enfermedad. En este caso, las redes neuronales convolucionales (CNN) son las técnicas empleadas actualmente para este fin, ya que en ciertas tareas llegan a superar la precisión de los expertos. El segundo caso correspondería a la clasificación de una parte de la imagen en dos o más clases, para lo cual son muy importantes tanto la información local de la lesión como el contexto a su alrededor.
- Detección de órganos o lesiones, más centrado en la localización, tanto espacial como temporal. Esta es una parte fundamental del diagnóstico y es de las más intensas para los médicos. Se ha trabajado mucho en los sistemas de detección por computadora que son diseñados para detectar lesiones automáticamente, mejorando notablemente la precisión de la detección, así como el tiempo invertido por los expertos.
- Segmentación de órganos y subestructuras, permitiendo la caracterización de parámetros relacionados con la forma o el volumen en análisis cardíacos o cerebrales, entre otros. La segmentación también se puede aplicar al caso de lesiones, combinando los retos de la detección de objetos, así como la segmentación de órganos y subestructuras aplicando algoritmos de deep learning. Actualmente, la segmentación es la temática más común en artículos sobre la aplicación del deep learning en imagen médica, por lo que parece lógico observar la gran variedad de metodologías estudiadas, incluyendo el desarrollo de arquitecturas CNN únicas o la aplicación de redes neuronales recurrentes (RNN), que cada vez están tomando más protagonismo en tareas de segmentación.
- Registro, o alineamiento espacial de imágenes médicas, una tarea muy común en el análisis de imágenes que se lleva a cabo para que todas las imágenes que conforman un dataset estén encuadradas respecto a la misma referencia. Esta técnica es empleada sobre todo cuando las

2.3. ALGORITMOS DE DEEP LEARNING PARA SEGMENTACIÓN DE TUMORES CEREBRALES

imágenes pueden desalinearse debido a movimientos de los pacientes, o a movimientos involuntarios de órganos, como los pulmones o el corazón.

- Otras tareas, como la búsqueda automática de imágenes de una temática dada dentro de una bases de datos masiva. Este proceso se realiza con la intención de identificar casos similares, enfermedades raras o incluso para mejorar los cuidados de los pacientes. Otras aplicaciones son la de la generación y mejora de imágenes a partir de arquitecturas CNN, cada vez aplicado a más tareas de deep learning, o la de elaboración de informes médicos a partir de una imagen.

2.3. Algoritmos de Deep Learning para segmentación de tumores cerebrales

Metodológicamente, muchos algoritmos de segmentación de tumores se basan en técnicas desarrolladas inicialmente para otras estructuras o patologías, como puede ser la segmentación automática de lesión de materia blanca, la cual posee una gran precisión.

En este ámbito de segmentación, los modelos de análisis de imagen más exitosos hoy en día son las redes neuronales convolucionales (CNN), en las que se lleva trabajando desde finales de los años 70, y que se empezaron a aplicar a imagen médica en 1995. Dentro de este tipo de redes, la arquitectura más conocida en análisis de imagen médica es probablemente la U-Net, comprendida por una CNN normal, de contracción y reducción de la dimensionalidad, seguida por una parte de expansión donde se usan deconvoluciones para aumentar la dimensionalidad de la imagen. La novedad que introduce la U-Net es la simetría, es decir, la combinación del mismo número de capas downsampling que de upsampling, así como el empleo de “skip connections” entre convoluciones y deconvoluciones opuestas. Este procedimiento concatena las características extraídas en el proceso de contracción con las capas de expansión.

En este tipo de redes, durante el proceso de contracción la información sobre características va aumentando mientras que la espacial se reduce. Por otra parte, el camino de expansión combina la información de características con la espacial mediante una secuencia de convoluciones inversas y de concatenaciones con características de alta resolución del camino de contracción [19]. En la sección 4.3 se proporcionarán más detalles acerca de este tipo de redes, puesto que ha sido el escogido para elaborar el modelo de segmentación de gliomas.

2.4. BraTS challenge

Para poder evaluar el actual estado del arte en cuanto a algoritmos de segmentación automática de gliomas y comparar los diferentes métodos existentes, en 2012 se empezó a organizar el challenge de BraTS (Multimodal Brain Tumor Image Segmentation Benchmark), en colaboración con MICCAI (Medical Image Computing and Computer Assisted Interventions).

BraTS siempre se ha centrado en la evaluación del estado del arte en cuanto a métodos para la segmentación de tumores cerebrales en escáneres de resonancia magnética multimodal. En concreto, el reto de 2018, del cual se han tomado los datos para llevar a cabo este trabajo, emplea secuencias de RM preoperatoria tomadas de diferentes instituciones, y se centra en la segmentación de gliomas. Por otro lado, BraTS también propone como segundo objetivo del reto la predicción de la tasa de supervivencia de los pacientes [18], pero este aspecto no ha desarrollado en el presente Trabajo de Fin de Master.

En el Capítulo 3 se describirá más detalladamente el dataset de imágenes de RM provisto por BraTS 2018, el cual se ha empleado para entrenar la red de segmentación automática de gliomas. Asimismo, se darán ejemplos de otras bases de datos de imagen médica de cerebro.

Capítulo 3

Dataset

En este Capítulo se da una visión genérica del estado del arte en cuanto a bases de datos de imagen médica, y se describe el conjunto de datos empleado en este trabajo.

3.1. Algunos repositorios de imagen de cerebro

El banco de datos BRAINS Imagebank (Brain Images of Normal Subjects) (<http://www.brainsimagebank.ac.uk>) es un proyecto para crear un repositorio integrado organizado por el Brain Research Imaging Centre de la Universidad Edimburgo, y financiado por los colaboradores de la SINAPSE (Scottish Imaging Network: A Platform for Scientific Excellence). BRAINS permite compartir y archivar imágenes detalladas del cerebro, previamente anonimizadas, así como los datos fenotípicos más relevantes, recolectados en estudios con individuos sanos de todas las edades. Particularmente se centra en los rangos más extremos, como ancianos y prenatales, donde la variabilidad es mayor, y los cuales además están infrarrepresentados en otras bases de datos existentes.

BRAINS es un banco de imágenes dinámico, el cual va incorporando nuevas imágenes según están disponibles. Actualmente consta con datos de 808 voluntarios sanos, con edades entre 15 y 81 años. Esta base de datos almacena diferentes secuencias de Imagen de Resonancia Magnética, como T1, T2 o FLAIR, disponibles en formato DICOM (<http://dicom.nema.org/>). Además, las imágenes están vinculadas a un amplio espectro de metadatos clínicos, tales como la edad, el historial médico, medidas fisiológicas, el uso de medicamentos, habilidades cognitivas o información del embarazo (en caso de sujetos prenatales y recién nacidos) [20].

Otros ejemplos de bases de datos de imagen médica son, por ejemplo, OA-SIS Brains (Open Access Series of Images Studiens) [21], TCIA (The Cancer Image Archive) [22], o ADNI (Alzheimer’s Disease Neuroimaging Initiative) [23].

3.2. Dataset BraTS 2018

Para la realización de este Trabajo de Fin de Máster, se ha empleado el conjunto de datos de imagen médica proporcionada por el Data Challenge BraTS 2018 (Brain Tumor Segmentation) [24].

Para poder obtener dichos datos, ha sido necesario registrarse en el Portal de Procesamiento de Imagen de CBICA [25], indicando el propósito para el cual desean emplearse los datos.

3.2.1. Descripción del dataset de entrenamiento

El conjunto de datos de entrenamiento se compone de 237 archivos, cada uno de los cuales es un estudio de Resonancia multimodal de un paciente distinto. En este caso, 162 corresponden a casos de pacientes con HGG (glioma de alto grado, III o IV), y los 75 restantes son casos de pacientes con LGG (glioma de bajo grado, I o II). En el caso de este proyecto, solo se han tenido en cuenta los 162 pacientes con HGG. Cada archivo corresponde a la imagen de Resonancia Magnética de un paciente distinto, y consta de:

- Secuencias ponderadas en T1.
- Secuencias ponderadas en T1, tras la administración de un contraste de gadolinio.
- Secuencias ponderadas en T2.
- Secuencias FLAIR (FLuid Attenuation Inversion Recovery, o recuperación por inversión atenuada de fluido), que son una prolongación de las T2, con supresión del líquido cefalorraquídeo.
- Imágenes segmentadas, en las cuales solamente se muestra el tumor, diferenciando las tres partes más relevantes: el tumor remarcado, el edema peritumoral, y la zona necrótica y no remarcada del núcleo del tumor.

Al tratarse de un volumen, se puede elegir el eje deseado para la visualización del cerebro. En este trabajo, como suele hacerse generalmente en otros

estudios, se ha tomado la vista axial del cerebro puesto que es la más común en el análisis de imagen médica, aunque cualquier otra vista hubiera sido válida para entrenar la red. Todas las imágenes que constituyen el dataset han sido segmentadas manualmente por varios expertos, que han seguido los mismos criterios de anotación. Además, todas ellas han sido proporcionadas con supresión del cráneo, de modo que en el preprocesamiento no ha sido necesaria su eliminación [24]. En la Figura 3.1 se muestra una vista sagital de cada tipo de secuencia: T1, T1 con contraste, T2, FLAIR e imagen segmentada.

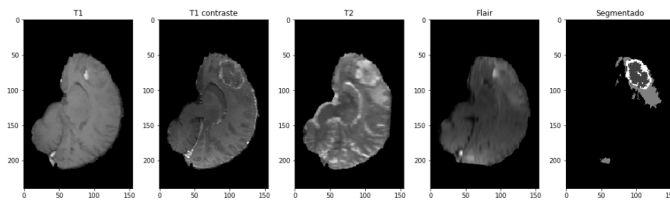


Figura 3.1: Diferentes secuencias de Resonancia Magnética del cerebro de un paciente. Para la misma sección de cerebro se muestran la T1, T1 con contraste, T2, flair y la imagen segmentada. Estos datos son procedentes del dataset de BraTS 2018 [24], y forman parte del conjunto que se ha empleado para entrenar la red. En la imagen segmentada se han tenido en cuenta las siguientes etiquetas para los píxeles: 1 para la parte necrótica y no resaltada, 2 para el edema peritumoral, 4 para el tumor resaltado, y 0 para el resto.

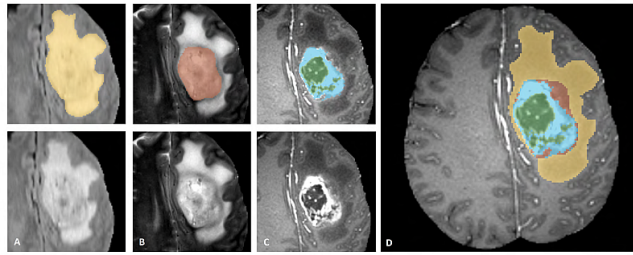


Figura 3.2: Segmentación manual de una imagen de RM realizada por los expertos en anotación. Arriba a la izquierda se muestran las estructuras que se pueden distinguir en las diferentes modalidades de imagen, siendo la imagen de la derecha la segmentación final de la imagen. De izquierda a derecha: tumor completo visible en secuencia FLAIR (A), el núcleo del tumor visible en T2 (B), las estructuras del tumor remarcado, en azul, rodeando a la parte necrótica, en verde, ambas visibles en la T1 con contraste (C). Estas componentes se combinan para generar la segmentación final de las estructuras del tumor (D). Fuente: [18].

Las regiones segmentadas consideradas para la evaluación del Data Challenge son las siguientes: 1) el “tumor resaltado” (ET), 2) el “core del tumor” (TC), y 3) el “tumor completo” (WT), ver Figura 3.2. Sin embargo, por simplicidad, en este trabajo se ha entrenado la red con etiquetas binarias, de modo que solo se identifica si el píxel corresponde o no a tumor.

El ET lo conforman las áreas que se muestran hiperintensas en secuencias T1 con contraste, al compararlas con la T1 normal, pero también al compararlo con materia blanca “sana” en la T1 con contraste. El TC representa el tumor en sí, que engloba el ET, así como la parte necrótica (con fluido) y las partes no resaltadas (sólidas). Estas partes suelen mostrarse hipointensas en la T1 con contraste, en comparación con la T1. Por último, el WT describe la lesión al completo, englobando las partes descritas anteriormente junto con el edema peritumoral (ED), el cual suele mostrarse hiperintenso en las secuencias FLAIR [26].

3.2.2. Descripción del dataset de validación

El dataset de validación consta de 66 archivos, uno por paciente, con los mismos modos que para las imágenes de entrenamiento, pero con la diferencia de que en este caso no se proporciona la imagen segmentada. La finalidad del Data Challenge BraTS 2018 es predecir, a partir de un modelo entrenado, cuál sería la segmentación de cada imagen, y enviar los resultados para evaluación por parte de un tribunal. Dicha evaluación consiste en comparar la predicción de imágenes segmentadas con la segmentación real.

Sin embargo, en este trabajo no se han enviado resultados de validación para que fueran evaluados por la comisión del Data Challenge de BraTS, puesto que la red se ha entrenado únicamente con etiquetas binarias que identifican la totalidad del tumor, sin distinguir subestructuras del mismo. Dado que las imágenes segmentadas originales diferencian tres partes del tumor, no tendría sentido enviar los resultados binarios para su comparación.

3.2.3. Formato de las imágenes

El formato de las imágenes proporcionadas por BraTS es NIfTI (.nii). NIfTI (Neuroimaging Informatics Technology Initiative) es una iniciativa impulsada por el Instituto Nacional de la Enfermedad Mental y el Instituto Nacional de Desórdenes Neurológicos y Derrame Cerebral, y cuyo principal objetivo es ofrecer los servicios necesarios para desarrollar herramientas informáticas relacionadas con neuroimagen [27].

Este formato permite almacenar tanto los datos de una imagen del cerebro en 3 dimensiones, como metadatos sobre la misma. La ventaja del formato .nii

respecto a otros, como DICOM, es que los metadatos no incluyen información sobre el paciente, lo cual es muy importante para preservar su privacidad.

Capítulo 4

Métodos

En este Capítulo se describen brevemente las librerías empleadas a la hora tanto de cargar las imágenes que han compuesto el dataset, como de construir la arquitectura del modelo. También se describen los diferentes pasos seguidos a lo largo de este trabajo.

4.1. Librerías de Python

Python ha sido el lenguaje de programación escogido para construir y entrenar el modelo de red neuronal, dada su gran versatilidad y amplio abanico de librerías específicas para este fin.

4.1.1. NiBabel

Esta librería soporta una creciente colección de formatos de archivos de neuroimagen. Como cada formato tiene sus propias características y peculiaridades, NiBaBel ofrece acceso independiente tanto a las imágenes como a la información que contiene dicho formato. Al cargar un imagen, NiBabel intenta reconocer su formato apartir del nombre del archivo

Este paquete es especialmente útil con el formato NIfTI, ya que permite leer fácilmente archivos de este tipo y analizar las imágenes directamente desde Python. Una vez cargada la información contenida en el archivo, NiBabel hace las imágenes accesibles mediante arrays de NumPy [28].

4.1.2. Keras

Se trata de una API de redes neuronales de alto nivel, escrito en Python y capaz de correr sobre TensorFlow, CNTK o Theano. Keras es una librería

de deep learning que permite elaborar fácilmente prototipos de redes neuronales, soportando tanto CNN (redes convolucionales) como RNN (redes recurrentes), y pudiendo ser ejecutada en CPU y GPU.

Keras es de fácil manejo, ya que está diseñado para humanos y no máquinas, permite crear nuevos modelos y modificar modelos ya existentes y, al estar escrito en Python, permite detectar errores más fácilmente que con otros lenguajes de programación [29].

4.2. Preprocesamiento de las imágenes

El código implementado para realizar el preprocesamiento de las imágenes se encuentra disponible en el siguiente enlace, en versión Notebook de Jupyter: https://github.com/andrea24396/TFM_Data_Science.

Lo primero que se ha realizado tras obtener los datos ha sido preprocesar las imágenes, de modo que se adecuasen al formato necesario para entrenar la red.

Uno de los primeros pasos ha sido reducir el número de datos que contiene el dataset. Para ello, puesto que se disponía de un gran número de imágenes de RM procedentes de 162 pacientes diferentes (teniendo solo en cuenta aquellos con HGG), se ha decidido prescindir de los datos de pacientes provenientes de la base de datos de TCIA, ya que al mostrar imágenes de este set se ha observado que la resolución es bastante baja.

Por otro lado, las imágenes que conforman las periferias del volumen escaneado son completamente negras, por lo que no aportan información relevante para la red y solo aumentan el volumen del dataset. Es por esto que parece lógico prescindir de ellas, para que el entrenamiento no se demore innecesariamente.

El procedimiento seguido ha sido el siguiente:

1. Guardar las imágenes de RM y sus etiquetas correspondientes en listas separadas.
2. Tras convertir estas listas en arrays de Numpy, se han eliminado las imágenes de los 56 pacientes provenientes de la base de datos de TCIA.
3. A continuación se ha realizado la separación en train y test, con las imágenes de los 81 primeros pacientes para train, y los restantes 27 para test.

4. Se ha eliminado la agrupación por pacientes, de modo que todas las imágenes forman un solo conjunto global.
5. Se han normalizado las imágenes. Para ello, se ha obtenido el valor máximo de los píxeles de cada imagen de RM, y se ha dividido cada una por su correspondiente valor máximo. De este modo se consigue que todas las imágenes sean mapas de píxeles con valores entre 0 y 1.
6. Se han eliminado las imágenes cuya suma de píxeles no supere un cierto umbral establecido. Esto implica que las imágenes completamente negras (suma = 0) o que representan muy poca superficie de cerebro (primeras rebanadas) han sido eliminadas del conjunto por no aportar información relevante para la red. Para poder establecer un umbral común para todas las imágenes, es imprescindible que estas hayan sido normalizadas previamente. Este punto se ilustra en la Fig. 4.1.

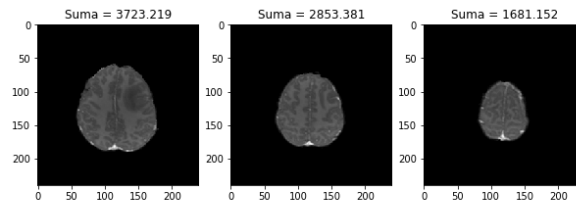


Figura 4.1: Tres rebanadas de cerebro casi consecutivas, encima de las cuales se muestra la suma total de píxeles.

7. Por último, se han binarizado las imágenes segmentadas. Como se describía en la Sección 3.2.1, las imágenes segmentadas provistas por el Challenge de BraTS (imagen del centro en Figura 4.2) tienen cuatro etiquetas diferentes, 0, 1, 2 y 4. Al binarizarlas (imagen de la derecha en Figura 4.2), se ha conseguido que los píxeles tengan como posibles valores 1 (tumor) o 0 (no tumor), sin hacer distinciones en cuanto a las diferentes estructuras del tumor.

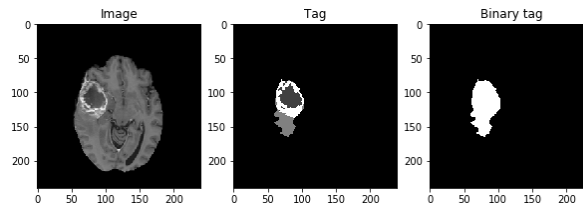


Figura 4.2: Proceso de binarización de la etiqueta de una imagen de RM.

4.3. Arquitectura de la red

Se ha tomado la U-net implementada en [30] como referencia para construir la red entrenada en este trabajo. Sin embargo, dada la reducida capacidad de procesamiento, no se ha podido alcanzar el nivel de complejidad de la misma. El código en formato Notebook, con la implementación del modelo, así como su entrenamiento y posterior obtención de resultados, se encuentra disponible en https://github.com/andrea24396/TFM_Data_Science

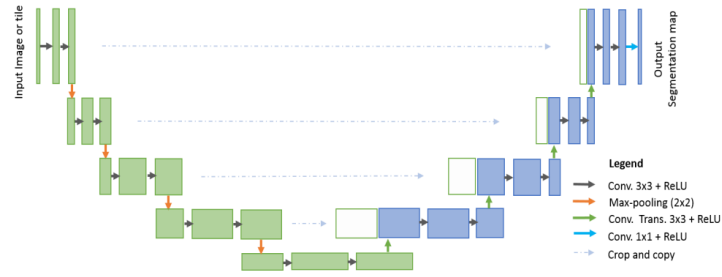


Figura 4.3: Arquitectura de U-Net.

La U-Net entrenada se compone de los siguientes elementos, ilustrados en la Figura 4.3:

- Camino de contracción, el cual está compuesto por tres bloques de dos capas convolucionales 3x3 con activación ReLU, seguidas de una capa de downsampling, en este caso maxpooling 2x2, y dropout. Las capas convolucionales del primer bloque tienen 16 filtros, número que se va doblando cada vez que se pasa al siguiente bloque. Por tanto, el camino de contracción consta de cuatro bloques con capas convolucionales de 16, 32, 64 y 128 filtros, respectivamente. Esta primera parte de la red tiene como objetivo reducir la dimensionalidad de las imágenes.
- Un bloque central, con dos capas convolucionales de 256 filtros 3x3 con activación ReLU.
- Camino de expansión, compuesto por otros tres bloques, simétricos a los anteriores, cada uno con una capa convolucional inversa 2x2 (up-convolution) seguida por dropout y dos capas convolucionales convencionales 3x3 con activación ReLU. La función de estos bloques es la de aumentar la dimensionalidad de las imágenes, hasta alcanzar las dimensiones de entrada. Además, como se explica en la sección 2.3, los niveles del camino de contracción se concatenan con los del camino de expansión para incluir la información extraída en el primero.

4.4. DEFINICIÓN DEL MODELO Y ENTRENAMIENTO DE LA RED²⁵

- Una capa final convolucional, de un filtro 1x1 con activación sigmoideal, para obtener el output de la misma dimensionalidad que el input. El output es, por tanto, el mapa de segmentación de la imagen introducida como input.

4.4. Definición del modelo y entrenamiento de la red

Una vez se ha definido la arquitectura de la red, se ha compilado y entrenado el modelo. Los parámetros que se han escogido para dichas etapas se muestran en la Tabla 4.1, y la partición de train y test considerada para entrenar los modelos se muestra en la Tabla 4.2.

Además, en el entrenamiento de los modelos se ha introducido un vector de pesos para compensar el desbalanceo de píxeles de tumor, 1, y píxeles de no tumor, 0 (en la Figura 4.2 puede observarse la gran descompensación que hay entre la parte de la imagen correspondiente al tumor sobre la parte de fondo). Si este desbalanceo no se corrigiera, la red tendería a predecir solo las clases con mayor representación (fondo negro), e ignorará las menos representativas (tumor blanco), por lo que no tendrían repercusión en el error total. De este modo, al introducir dichos pesos, se espera que los píxeles de tumor puedan estar igualmente representados que los del fondo.

El entrenamiento se ha realizado a través de una GPU en Jupyter Lab, dado que en una CPU el entrenamiento es mucho más lento y con menos capacidad de procesamiento. Por ejemplo, los primeros entrenamientos realizados en CPU tardaban entre 12 y 20 horas para completar 6 épocas, mientras que con la GPU los mismos modelos tardaron alrededor de 20 minutos para completar 20 épocas.

Compile		Fit	
Optimizador	Adam (lr = 0.001)	Épocas	20
Pérdida	Binary crossentropy	Early stopping	patience = 5
Métrica	Accuracy	Batch size	20
		Validation split	0.2

Tabla 4.1: Parámetros que configuran la compilación y el entrenamiento del modelo creado.

	T1	T1ce
Train	7394	4725
Test	2454	1590
Total	9848	6315

Tabla 4.2: Partición train-test de los conjuntos de imágenes considerados para entrenar el modelo. Se han considerado secuencias T1 y T1 con contraste, indicado como T1ce.

4.5. Predicciones

Una vez completado el entrenamiento, se ha realizado la predicción sobre el conjunto de test, obteniendo así la segmentación de las imágenes pasadas a la red. El output obtenido en la predicción se corresponde con los mapas de segmentación del input, y se han comparado con las segmentaciones reales para evaluar la calidad de la predicción.

Para ello se ha considerado apropiado calcular el coeficiente DICE, que evalúa la similitud entre la etiqueta predicha y la considerada como verdadera, “ground truth”. Este se calcula mediante la Ec. (4.1):

$$\text{DICE} = \frac{2 * (A \cap B)}{|A| + |B|} \quad (4.1)$$

donde A y B representan la predicción y la imagen segmentada original. Para obtener un valor más fiel a la realidad, la etiqueta predicha se ha binarizado, de modo que los píxeles solo pudieran tener valores 0 o 1 en base a un umbral dado, y así poder comparar mejor predicción y ground truth.

Dado el gran desbalanceo de píxeles con valores 0 o 1 en las imágenes segmentadas y binarias, este coeficiente es bastante apropiado para cuantificar la calidad de la predicción, ya que se centrará en los píxeles que conforman el tumor. Por otro lado, una desventaja es que incluso obteniendo una buena predicción, en cuanto a la posición y forma aproximada del tumor, el coeficiente DICE puede verse perjudicado si la forma del tumor predicho no se asemeja al original. Este aspecto se discutirá en el Capítulo 6.

Capítulo 5

Resultados y análisis

En este Capítulo se presentan los resultados obtenidos para la segmentación de gliomas con dos modelos diferentes.

5.1. Introducción

En las siguientes secciones se hará referencia a dos conjuntos de datos proporcionados por BraTS, Training y Validation. Los primeros son los empleados para entrenar los modelos, dividiéndolos previamente en train y test, y que se describen en la sección 3.2.1, mientras que los segundos son los descritos en la sección 3.2.2, y son los empleados en la validación de los modelos.

Por otro lado, solo se han empleado imágenes T1 y T1 con contraste para entrenar los modelos, porque se considera que estas dos secuencias son las que aportan la mejor información en cuanto al tumor, y se ha querido comparar la capacidad de predicción con cada una de ellas.

Tras haber probado inicialmente varias arquitecturas de U-net, con distinto número de niveles de profundidad y diferentes parámetros de entrenamiento, se ha tomado como definitivo el modelo descrito en la sección 4.3. Con este modelo se han hecho predicciones sobre los conjuntos de test y validation, obteniendo la segmentación automática para imágenes que la red no había visto nunca previamente.

5.2. Resultados con datos de Training

5.2.1. Primeros resultados

En un primer momento, se entrenaron dos modelos distintos a partir de la misma arquitectura, descrita en la sección 4.3, uno con imágenes T1 y otro con imágenes T1 con contraste. Una vez completados ambos entrenamientos, se hicieron las predicciones sobre los correspondientes conjuntos de test, con la finalidad de obtener la segmentación de las imágenes. En el Apéndice A se da una muestra de algunas predicciones para imágenes T1.

Para las segmentaciones obtenidas, binarizadas con 0.5 como umbral, se calculó el coeficiente DICE, obteniendo la media y desviación estándar del conjunto:

$$\text{DICE_T1} = 0.54 \pm 0.42$$

$$\text{DICE_T1ce} = 0.53 \pm 0.41$$

A pesar de los resultados tan bajos del coeficiente DICE, se observó que las predicciones eran cualitativamente bastante buenas en cuanto a la forma y posición del tumor. Sin embargo, se pensó que este buen resultado se obtuvo debido a que en el preprocesamiento del dataset la partición train-test se hizo al final del mismo, una vez eliminada la división de imágenes por pacientes y aleatorizadas en el conjunto global. Esto hubiera supuesto que al realizar la partición 75-25 sobre este conjunto global, habría imágenes consecutivas del mismo paciente en ambos conjuntos. Esta situación se traduce en que una imagen de test tendría otra prácticamente idéntica en train, y hubiera sido igual que predecir sobre una imagen que ha intervenido en el entrenamiento.

En consecuencia, se modificó el pipeline del preprocesamiento, haciendo la partición train-test al comienzo del proceso, esta vez respecto a los pacientes, y se volvieron a entrenar dos modelos, uno con imágenes T1 y el otro con T1 con contraste.

5.2.2. Resultados definitivos

Una vez modificado el preprocesamiento de las imágenes, realizando la partición train-test al comienzo, se disponía de dos conjuntos de imágenes totalmente independientes, cada uno de ellos con imágenes de pacientes diferentes. Por tanto, se entrenaron de nuevo los modelos con los conjuntos de train, y se hicieron las predicciones sobre los conjuntos de test. De este modo se obtuvieron unos nuevos resultados, de los cuales se presenta una pequeña

muestra en el Apéndice B.

Para cuantificar la calidad de estas predicciones, se ha calculado el coeficiente DICE, como se describe en la sección 4.5. Primeramente se ha calculado sin binarizar las segmentaciones predichas, obteniendo $\text{DICE} = 0.02 \pm 0.03$. A continuación se han binarizado las imágenes predichas con distintos valores de umbral, observando que con cualquier valor se obtienen DICEs muy similares entre sí. Por tanto, se ha escogido 0.5 como umbral óptimo y se ha calculado el DICE para las predicciones sobre imágenes T1 y T1ce, obteniendo respectivamente:

$$\text{DICE_T1} = 0.64 \pm 0.39$$

$$\text{DICE_T1ce} = 0.57 \pm 0.41$$

5.3. Resultados con datos de Validation

Por último, se han probado los modelos con algunas imágenes del conjunto de Validation proporcionado por BraTS, destinado a la obtención de resultados y su posterior envío a la organización del Data Challenge de BraTS 2019 para su evaluación. Sin embargo, los resultados obtenidos en este trabajo no se han enviado a evaluar puesto que los modelos entrenados solamente predicen la totalidad del tumor, sin diferenciar las subestructuras del mismo.

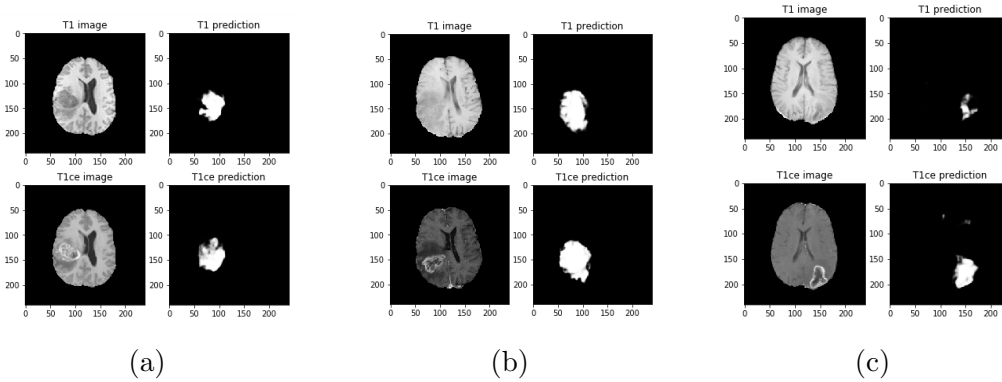


Figura 5.1: Muestra de algunas predicciones obtenidas con imágenes del conjunto de Validation.

Capítulo 6

Discusión

En el presente proyecto se ha estudiado la capacidad de segmentación de gliomas de alto grado, HGG, de un modelo de Deep Learning cuya arquitectura se basa en el empleo de redes neuronales convolucionales.

Primeramente se discutirán los resultados obtenidos para las predicciones con las imágenes de test.

A la vista de los primeros resultados obtenidos, se pensó que el hecho de que a simple vista fueran muy buenos se debía a que la partición de train-test se hizo al final del preprocesamiento, cuando las imágenes ya estaban distribuidas aleatoriamente en el conjunto global. En este momento, el coeficiente DICE obtenido para las predicciones sobre el conjunto de test era de 0.54 ± 0.42 para imágenes T1 y 0.53 ± 0.41 para T1 con contraste, donde el error se ha calculado como la desviación estándar de todos los coeficientes DICE calculados para cada imagen. Los DICE obtenidos, más bajos de lo esperado, así como su elevado error dan una visión de la gran variabilidad de acierto en cuanto a las predicciones. Esto se debe a que tanto los falsos positivos como los falsos negativos obtienen DICEs muy bajos, perjudicando a la tendencia general.

A continuación se cambió el procedimiento seguido en el preprocesamiento, y se hizo dicha división al comienzo, con el 75 % de los pacientes para train y el 25 % restante para test. Se observó que, aun así, los resultados obtenidos seguían siendo bastante buenos cualitativamente en cuanto a la forma y la localización espacial, mientras que los DICE calculados se mantenían de nuevo más bajos de lo esperado, 0.64 ± 0.39 para imágenes T1 y 0.57 ± 0.41 para T1 con contraste. De todos modos, estos resultados son mejores que los obtenidos en la primera fase del estudio.

Comparando ambas fases del estudio, se puede observar que en la Figura A.1b (primera fase), cuando la imagen input no tiene tumor, el modelo predecía una figura similar a la imagen pasada a la red. Esto puede deberse a las “skip connections” de la red, que de alguna manera mantienen información sobre el input. Por otro lado, en las Figuras B.1 y B.2 (segunda fase), cuando la imagen input no tiene tumor, la predicción resultante muestra un sombreado sin una forma determinada.

A continuación se discutirá el efecto de binarizar las segmentaciones obtenidas con el modelo.

Como se ha comentado en el capítulo anterior, es muy importante que las segmentaciones predichas sean mapas binarios para poder calcular un coeficiente DICE fiel a la realidad; en caso contrario cada píxel tendría un valor entre 0 y 1, y al ser comparado con un valor estrictamente 0 o 1 de la segmentación original, se obtendría un DICE muy bajo e irreal.

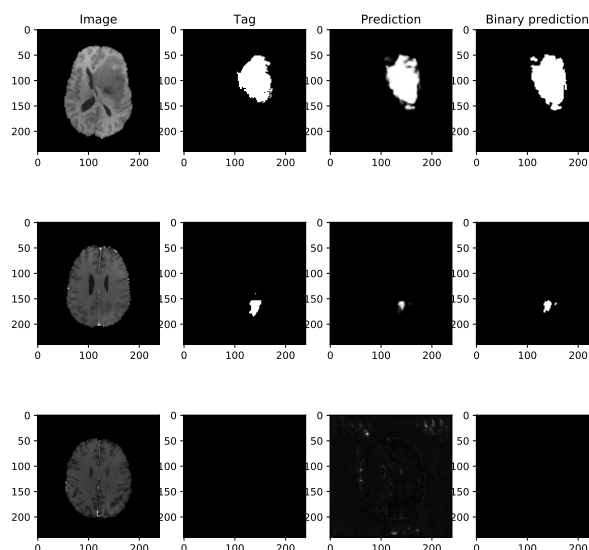


Figura 6.1: Resultado de binarizar las segmentaciones predichas. Se muestran tres imágenes al azar del conjunto de test. De izquierda a derecha, la imagen, la segmentación real, la predicción del modelo para esa imagen y la imagen segmentada binarizada respecto a un cierto umbral, en este caso, 0.1.

El proceso de binarización tras obtener las predicciones puede hacer que los resultados mejoren, aunque también puede surtir el efecto contrario. Esto se ilustra en la Figura 6.1.

En el caso central, las regiones que originalmente tienen valores de pixel más cercanos a 0 en la periferia del tumor predicho, pero superiores al umbral, pasan a ser blancas, haciendo que la predicción se asemeje más a la etiqueta real.

Pasa lo mismo en el último caso de dicha Figura, donde la predicción sobre una imagen que no presenta tumor genera algunas formas extrañas, llamadas artefactos, con valores mayores que 0, pero no lo suficiente elevadas para superar el umbral. Por tanto, al binarizarlas, el resultado será una imagen completamente negra.

Por último, en el primer caso pasa lo contrario, viéndose que la superficie que representa el tumor predicho es mayor que la etiqueta real después de binarizarla. Este sería un ejemplo de que binarizar puede a veces empeorar los resultados.

Para finalizar, se discutirán los resultados obtenidos con las imágenes de validación.

Las imágenes del conjunto de Validation han sido proporcionadas sin su “ground truth”, por lo que no hay posibilidad de cuantificar la calidad de las predicciones realizadas sobre ellas.

Sin embargo, sí se ha podido hacer una valoración cualitativa de las predicciones, escogiendo una serie de imágenes en las que a simple vista se puede diferenciar el glioma, ilustrado en la Figura 5.1. Así puede decirse que las predicciones son bastante buenas, ya que con cualquiera de los dos modelos, el de imágenes T1 o T1 con contraste, tanto la predicción de la posición como la de la forma del tumor es muy buena. Esta última característica no es determinada de igual forma por ambos modelos, pero se puede ver que en esencia describen un volumen similar.

Capítulo 7

Conclusiones y trabajo futuro

En este Capítulo se exponen las conclusiones obtenidas de la realización del presente proyecto así como futuras líneas de trabajo para continuar y mejorar el mismo.

7.1. Conclusiones

Se ha elaborado un modelo de Deep Learning basado en el empleo de redes neuronales convolucionales para segmentar gliomas en imágenes de Resonancia Magnética multimodal. Los tumores son segmentados en su totalidad sin diferenciar ninguna subestructura, a diferencia de otros modelos del Estado del Arte actual [18] en los que sí se diferencian partes. A la vista de los resultados obtenidos, y teniendo en cuenta que se trata de una arquitectura bastante sencilla, con unos parámetros de configuración del entrenamiento bastante genéricos, se pueden obtener las siguientes conclusiones:

- La etapa de preprocesamiento de las imágenes es un proceso clave que requiere de una atención especial y un estudio meticuloso, para garantizar la correcta preparación de las imágenes y lograr un entrenamiento satisfactorio.
- El objetivo principal de este trabajo ha sido conseguido, logrando que al pasarle al modelo una secuencia de RM de cerebro, la predicción obtenida sea la segmentación del tumor sobre un fondo negro.
- Los modelos diseñados, entrenados independientemente con imágenes T1 y T1 con contraste, predicen satisfactoriamente tanto la posición como la forma aproximada del tumor.

- La binarización de las segmentaciones predichas es un paso clave a la hora de calcular un DICE realista.
- Se obtienen mejores resultados cuando los conjuntos de train y test se definen en base a los pacientes, en lugar de hacerlo tras aleatorizar el conjunto global de imágenes.
- Sin embargo, los resultados obtenidos del coeficiente DICE son en general algo más bajos de lo esperado, debido probablemente a que los falsos positivos y falsos negativos influyen muy negativamente en el DICE medio.
- Resulta complicado contextualizar los resultados obtenidos para poder compararlos con los reportados en el estado del arte, debido a que el modelo diseñado realiza predicciones binarias, en lugar de multicategoría.
- Con todo esto, el método de segmentación desarrollado en este trabajo, se presenta como una herramienta útil a la hora de segmentar gliomas de alto grado.
- Los modelos basados en CNN, más en concreto aquellos con una arquitecturas simétricas de tipo U-Net, poseen un gran potencial a la hora de superar los primeros pasos de la segmentación de gliomas. Los resultados de las mismas, en cuanto a la determinación de la posición y forma aproximada, pueden llegar a compararse con las tareas realizadas manualmente por los expertos, pero reduciendo considerablemente el tiempo empleado para llevarlo a cabo.

7.2. Trabajo futuro

En cuanto a mejoras a corto plazo del modelo presentado en este trabajo, se propone incluir la definición de un parámetro que permita determinar la capacidad del modelo para localizar espacialmente los tumores. Como se ha observado cualitativamente en el Capítulo 5, el modelo localiza espacialmente muy bien los tumores. Una forma de cuantificarlo sería obtener el “centro de masas” del tumor en la segmentación original, y compararlo con el “centro de masas” del tumor predicho. Convirtiendo las posiciones a escala real (1 píxel cuadrado es aproximadamente 1 cm cuadrado), se podría obtener el error de localización espacial.

Otra mejora a considerar, siempre y cuando la capacidad de procesamiento de la máquina lo permitiera, sería entrenar el modelo con todas las

imágenes disponibles, para que el aporte de información sea mayor y más variado. Se podría también definir el modelo de forma que el input tuviera 4 canales en vez de 1, cada uno de los cuales se correspondería con cada una de las secuencias posibles de RM , T1, T1ce, T2 y FLAIR. De este modo se conseguiría que el tumor tenga el máximo aporte de información para su caracterización en la red.

A largo plazo, como líneas futuras se plantea entrenar el modelo con las etiquetas completas originales, en las cuales se diferencian varias regiones del tumor. De este modo, tanto el entrenamiento como la predicción pasarían a ser multicategoría en lugar de binarios.

También se podrían incluir imágenes de pacientes diagnosticados con LGG en el conjunto de entrenamiento, de modo que el modelo tenga un rango de aplicación más amplio en cuanto a pacientes con un posible tumor cerebral, pudiendo incluso añadir un factor de clasificación de tumor LGG o HGG.

Por último, como alternativa para construir la arquitectura de la red, se propone probar con una red basada en Redes Residuales, como se describe en [31].

Bibliografía

- [1] Pedro Pérez Segura. Tumores cerebrales. En: Sociedad Española de Oncología Médica (SEOM). (Disponible en: <https://seom.org/info-sobre-el-cancer/tumores-cerebrales>), Febrero 2017. [en línea] acceso 14-07-2019, 24-08-2019.
- [2] Neil M. Borden, Scott E. Forseen, and Cristian Stefan. *Imaging Anatomy of the Human Brain: A Comprehensive Atlas Including Adjacent Structures*. Springer Publishing Company, 2015.
- [3] MA Arraez, I Herruzo, T Acha, and M Benavides. Tumores del sistema nervioso central en el adulto y en la infancia. enfoque multidisciplinar neurooncológico. *Nova Sidonia Oncología y Hematología*. Madrid, 2012.
- [4] Wikipedia. La Enciclopedia Libre. Célula glial. (Disponible en: https://es.wikipedia.org/wiki/Célula_glial), 2019. [en línea] acceso 25-05-2019.
- [5] Asociación Española Contra el Cáncer (AECC). ¿qué son los tumores del sistema nervioso central? (Disponible en: <https://www.aecc.es/es/todo-sobre-cancer/tipos-cancer/sistema-nervioso-central/que-son-tumores-sistema-nervioso-central>). [en línea] acceso 25-08-2019.
- [6] Grupo Español de Investigación en Neurooncología (GEINO). Tipos histológicos. (Disponible en: https://geino.es/?page_id=205). [en línea] acceso 24-08-2019.
- [7] Alfredo Torregrosa LLoret. Diseño de una segmentación automática de gliomas en imágenes de resonancia magnética multimodal con redes neurales convolucionales profundas. 2018.
- [8] Giulio Metro. Glioma: guía para los pacientes - información basada en las directrices de práctica clínica esmo. [en línea] acceso 26-08-2019.

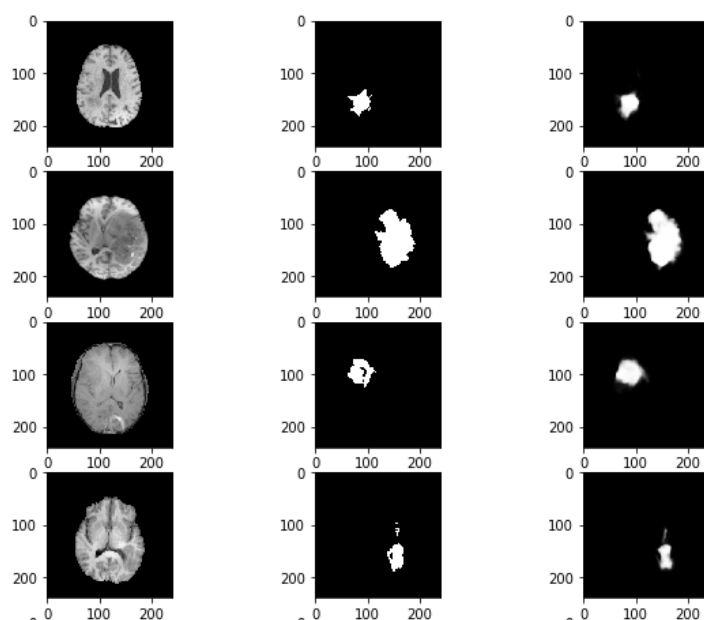
- [9] Iván Peñuelas. Radiofármacos pet. *Rev. Esp. Med. Nuclear*, 20(6):(477–498, 2001.
- [10] Dirk Pauleit et al. Comparison of o-(2-18f-fluoroethyl)-l-tyrosine pet and 3-123i-iodo- α -methyl-l-tyrosine spect in brain tumors. *Journal of Nuclear Medicine*, 45(3):374–381, 2004.
- [11] Kurt A Schalper, Maria E Rodriguez-Ruiz, Ricardo Diez-Valle, Alvaro López-Janeiro, Angelo Porciuncula, Miguel A Idoate, Susana Inogés, Carlos De Andrea, Ascensión López-Díaz de Cerio, Sonia Tejada, et al. Neoadjuvant nivolumab modifies the tumor immune microenvironment in resectable glioblastoma. *Nature medicine*, 25(3):470, 2019.
- [12] NIBIB. Instituto Nacional de Bioingeniería e Imágenes Biomédicas. Imagen por Resonancia Magnética (IRM). (Disponible en: <https://www.nibib.nih.gov/espanol/temas-cientificos/imagen-por-resonancia-magn%C3%A9tica-irm>), 2017. [en línea] acceso 22-08-2019.
- [13] Nelly Joseph-Mathurin et al. Magnetic resonance imaging in primates. the example of the mouse lemur (*microcebus murinus*): From detection of pathological aging to therapeutic evaluations. (Disponible en: <https://journals.openedition.org/primatologie/508>), 2010. [en línea] acceso 22-08-2019.
- [14] Sociedad Española de Imagen Cardíaca. ¿Qué es el T1 y el T2? En: Preguntas al radiólogo - Técnica y Metodología en Resonancia Magnética. (Disponible en: <https://ecocardio.com/documentos/biblioteca-preguntas-basicas/preguntas-al-radiologo/914-que-es-t1-y-t2.html>). [en línea] acceso 10-07-2019.
- [15] Damián Gil Bello. Semiología básica de la resonancia magnética. (UDIAT Centre Diagnòstic). (Disponible en: <http://www.acmcb.es/files/425-9513-DOCUMENT/Gil13Gen16.pdf>). [en línea] acceso 20-07-2019.
- [16] Sociedad Española de Imagen Cardíaca. ¿Por qué se emplea gadolinio como contraste en la RM? en: Preguntas al radiólogo - Algo de contrastes. (Disponible en: <https://ecocardio.com/documentos/biblioteca-preguntas-basicas/preguntas-al-radiologo/957-por-que-se-empieza-gadolinio-contraste-rm.html>). [en línea] acceso 10-07-2019.

- [17] Section for Biomedical Image Analysis (SBIA). Multimodal brain tumor segmentation challenge 2018. En: MICCAI BraTS 2018: Clinical Relevance. (Disponible en: <https://www.med.upenn.edu/sbia/brats2018/clinical-relevance.html>). [en línea] acceso 10-07-2019.
- [18] Bjoern H. MENZE et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34:1993–2024, 2014.
- [19] Geert Litjens et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.
- [20] Dominic E Job et al. A brain imaging repository of normal structural mri across the life course: Brain images of normal subjects (brains). *NeuroImage*, 144:299–304, 2017.
- [21] Open Access Series of Imaging Studies (OASIS). (Disponible en: <https://http://www.oasis-brains.org/>). [en línea] acceso 10-10-2019.
- [22] Kenneth Clark et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.
- [23] Alzheimer’s Disease Neuroimaging Initiative (ADNI). Multimodal brain tumor segmentation challenge 2018. En: MICCAI BraTS 2018: Clinical Relevance. (Disponible en: <https://http://adni.loni.usc.edu/>). [en línea] acceso 10-10-2019.
- [24] Section for Biomedical Image Analysis (SBIA). Multimodal brain tumor segmentation challenge 2018. En: MICCAI BraTS 2018: Data. (Disponible en: <https://www.med.upenn.edu/sbia/brats2018/data.html>). [en línea] acceso 10-07-2019.
- [25] Center for Biomedical Image Computing and Analytics, University of Pennsylvania. CBICA Image Processing Portal. A web accessible platform for imaging analitics. (Disponible en: <https://ipp.cbica.upenn.edu/>). [en línea] acceso 10-07-2019.
- [26] Section for Biomedical Image Analysis (SBIA). Multimodal brain tumor segmentation challenge 2018. En: MICCAI BraTS 2018: Tasks. (Disponible en: <https://www.med.upenn.edu/sbia/brats2018/tasks.html>). [en línea] acceso 10-07-2019.

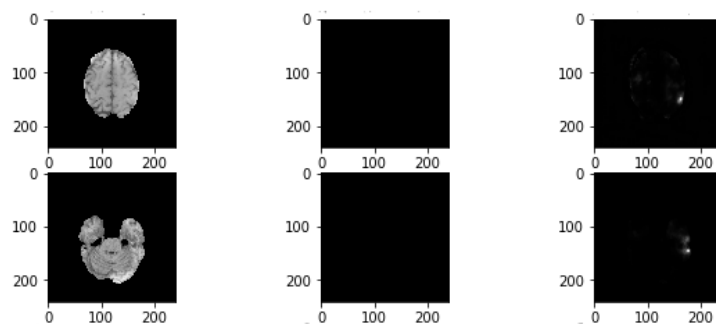
- [27] NIfTI: Neuroimaging Informatics Technology Initiative. (Disponible en: <https://nifti.nimh.nih.gov/>). [en línea] acceso 08-08-2019.
- [28] NiBabel: Access a cacophony of neuro-imaging file formats. (Disponible en: <https://nipy.org/nibabel/gettingstarted.html/>).
- [29] Keras: The Python Deep Learning Library. (Disponible en: <https://keras.io/>). [en línea] acceso 15-07-2019.
- [30] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).
- [31] Md Zahangir Alom et al. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. 2018.

Apéndice A

Primeros resultados



(a) Predicciones realizadas con el modelo entrenado con imágenes T1, para secciones de cerebro que presenta tumor.



(b) Predicciones realizadas con el modelo entrenado con imágenes T1, para secciones de cerebro que no presentan tumor. Puede observarse que el modelo predice “algo” cuando no debería haberlo.

Figura A.1: Muestra de las primeras predicciones obtenidas. En ambas imágenes, La columna de la izquierda muestra la imagen del cerebro considerada, la central se corresponde con la etiqueta real, el “ground truth”, y la de la derecha es la predicción obtenida.

Apéndice B

Resultados definitivos

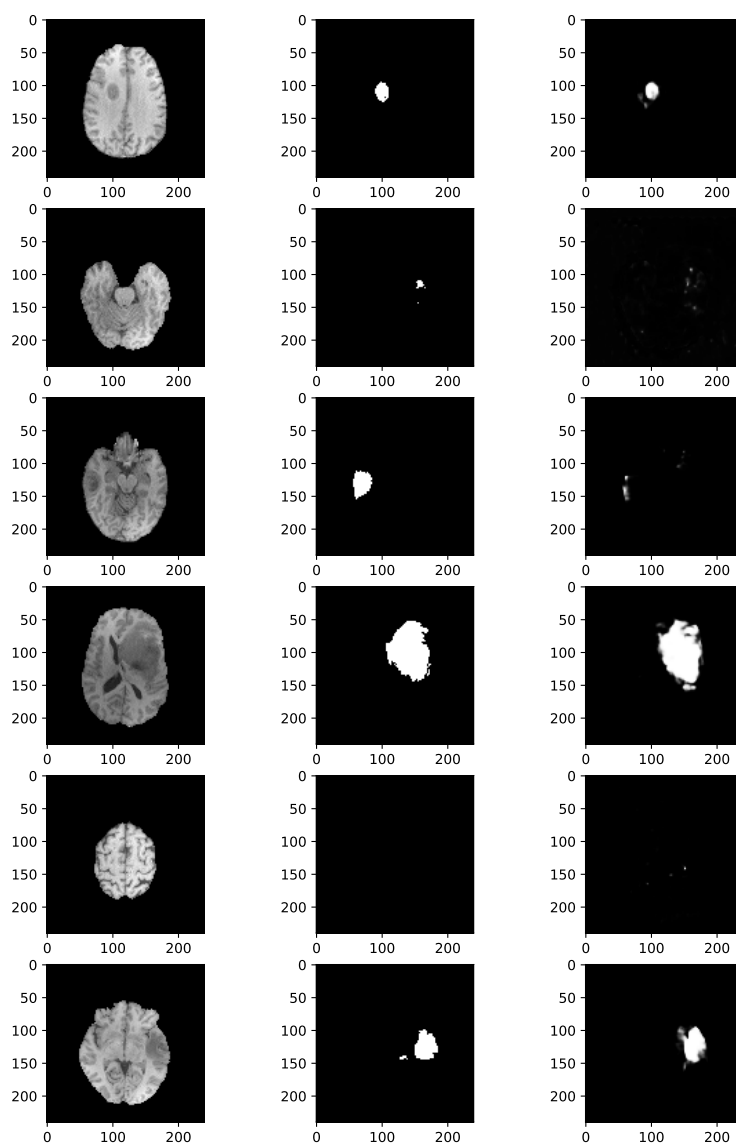


Figura B.1: Predicciones realizadas con el modelo entrenado con imágenes T1.

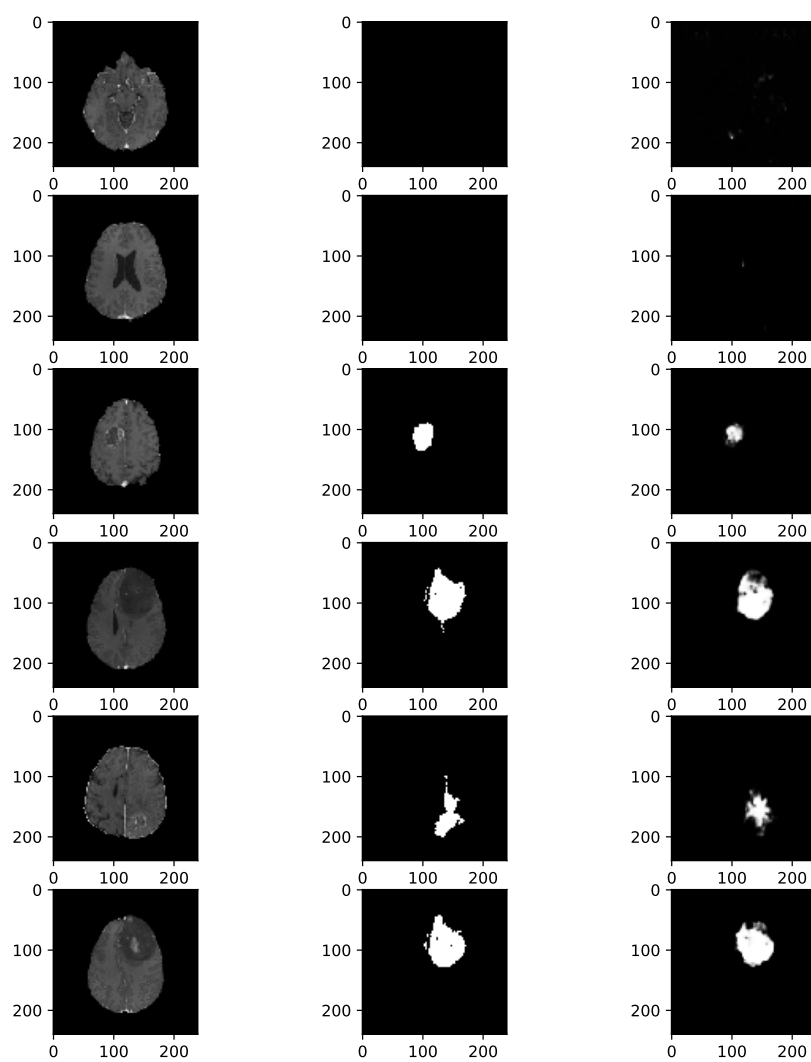


Figura B.2: Predicciones realizadas con el modelo entrenado con imágenes T1 con contraste.